High Throughput Exposure Assessment for Thousands of Chemicals

John Wambaugh
National Center for Computational Toxicology
Office of Research and Development
U.S. Environmental Protection Agency
wambaugh.john@epa.gov

Fit-For-Purpose Exposure Assessments
For Risk-Based Decision Making
Como, Italy
June 21-22, 2017

ORCID: 0000-0002-4024-534X

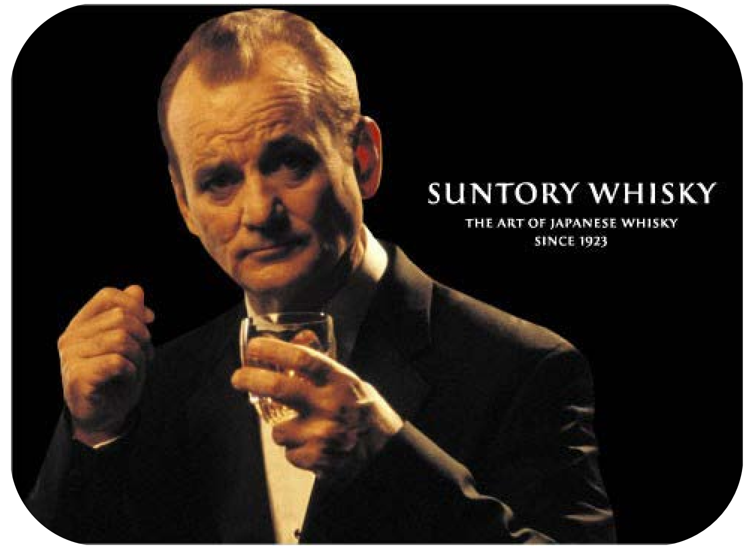
Fit for Purpose Models

Models Incorporate Knowledge, Assumptions and Data

- Training sets
- Choices of parameters
- Description of kinetics

A "fit for purpose" model is an abstraction of a complicated problem that allows us to reach a decision.

A fit for purpose model is defined as much by what is omitted as what is included in the model.

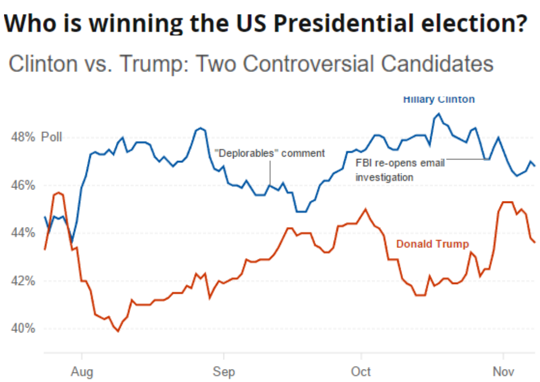


"The more you know who you are, and what you want, the less you let things upset you."
Bob, Lost in Translation via Todd Gouin
(Written by Sofia Coppola)

We have to accept that there will always be areas in need of better data and models -- our knowledge will always be incomplete, and thus we wish to extrapolate.

Parsimony and “Domain of Applicability”

- Do not build beyond the ability to evaluate predictions
- Collect data to allow larger, systematic studies
- Carefully determine whether, when, and why model errors are conservative and **correlated**



Daily Mirror/RealClearPolitics

Uncertainty Analysis on November 4:

FiveThirtyEight

Politics Sports Science & Health Economics Culture

NOV. 4, 2016 AT 11:09 AM

Trump Is Just A Normal Polling Error Behind Clinton

By Harry Enten

Filed under 2016 Election

How this was viewed at the time (November 5):

POLITICS 11/05/2016 03:59 pm ET | Updated Nov 07, 2016

Nate Silver Is Unskewing Polls — All Of Them — In Trump’s Direction

The vaunted 538 election forecaster is putting his thumb on the scales.

By Ryan Grim

*Model errors, especially **correlated** errors, matter*

Using 21st Century Science to Improve Risk-Related Evaluations

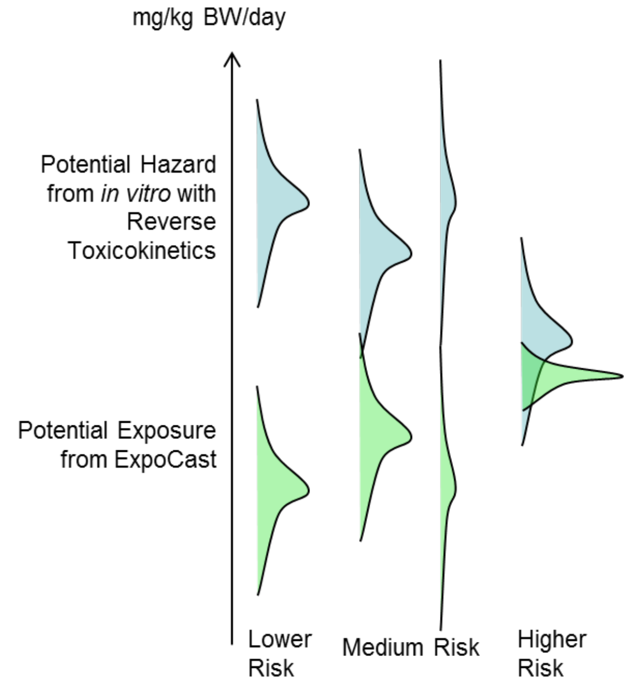
- **January, 2017 U.S. National Academies of Science report:**

  "Translation of high-throughput data into risk-based rankings is an important application of exposure data for chemical priority-setting. Recent advances in high-throughput toxicity assessment, notably the ToxCast and Tox21 programs… and in high-throughput computational exposure assessment… have enabled first-tier risk-based rankings of chemicals on the basis of margins of exposure"

- **Tox21/ToxCast**: Examining thousands of chemicals using in vitro assays that test parent chemical in concentration response

- **ExpoCast**: Tentative exposure predictions for daily human exposure rates (mg/kg/day)

- What is acceptable uncertainty?

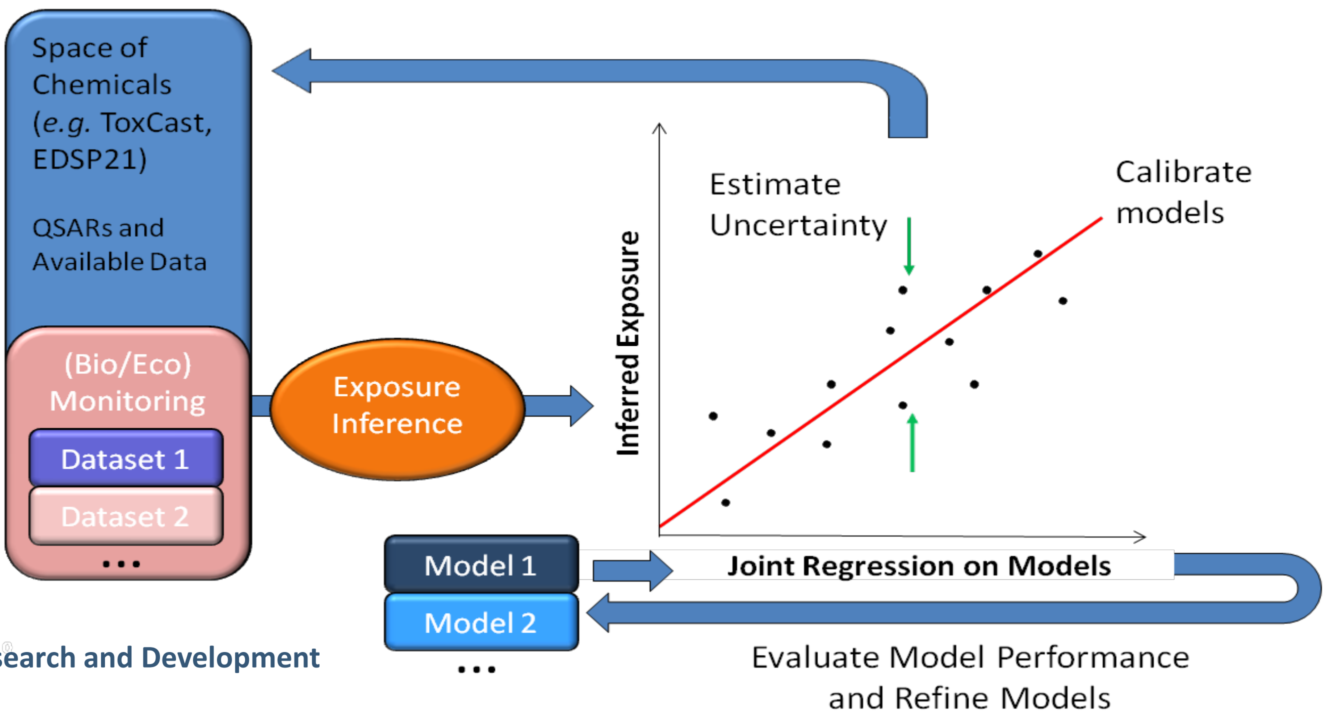


A fit for purpose exposure model might provide context for high throughput *in vitro* toxicity screening

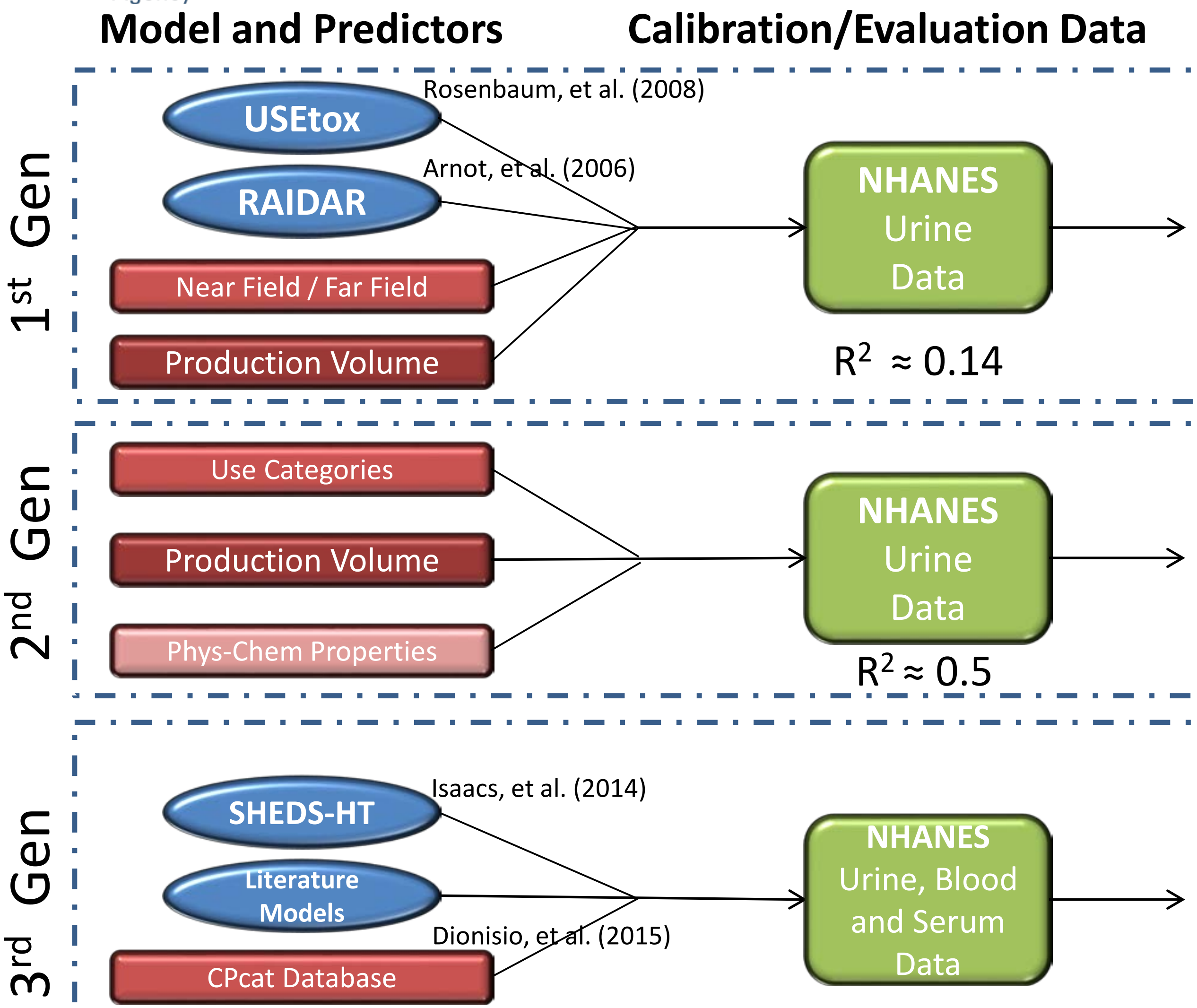
See Wetmore et al. (2015)

Consensus Exposure Predictions with the SEEM Framework

- We incorporate multiple models into consensus predictions for 1000s of chemicals within the **Systematic Empirical Evaluation of Models (SEEM) framework** (Wambaugh et al., 2013, 2014)
- We evaluate/calibrate predictions with available monitoring data across as many chemical classes as possible to allow extrapolation
  - Attempt to identify correlations and errors empirically



SEEM Evolution



SEEM Conclusion

- Existing complex fate and transport models have low correlation to measured exposures
- Near field factor most important

Wambaugh et al. (2013)

- Simple, readily available data
- Better correlation to measured exposures
- Similar predictions across demographics

Wambaugh et al. (2014)

- Need volume of distribution predictions (httk package) to use NHANES blood and serum data
- Analysis is ongoing

Ring et al. (*in prep.*)

Approach described in Wambaugh *et al.* (2013) ExpoCast Framework Paper

Heuristics of Exposure

Wambaugh et al. (2014)

Five descriptors explain roughly 50% of the chemical to chemical variability in median NHANES exposure rates

Same five predictors work for all NHANES demographic groups analyzed – stratified by age, sex, and body-mass index:

- Industrial and Consumer use
- Pesticide Inert
- Pesticide Active
- Industrial but no Consumer use
- Production Volume

Exposure Pathways

United States Environmental Protection Agency

Why does a heuristic model like Wambaugh et al. (2014) work?

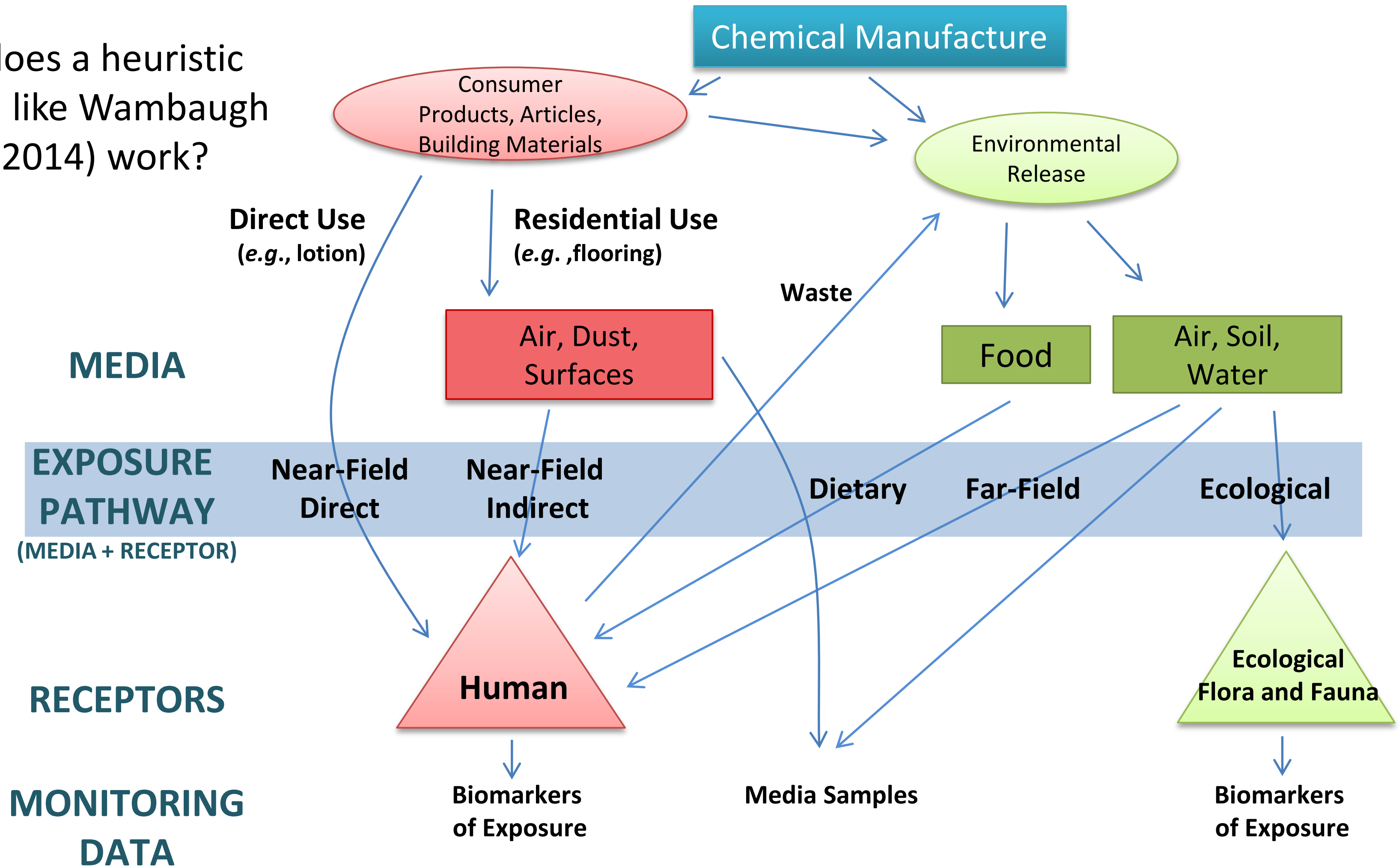
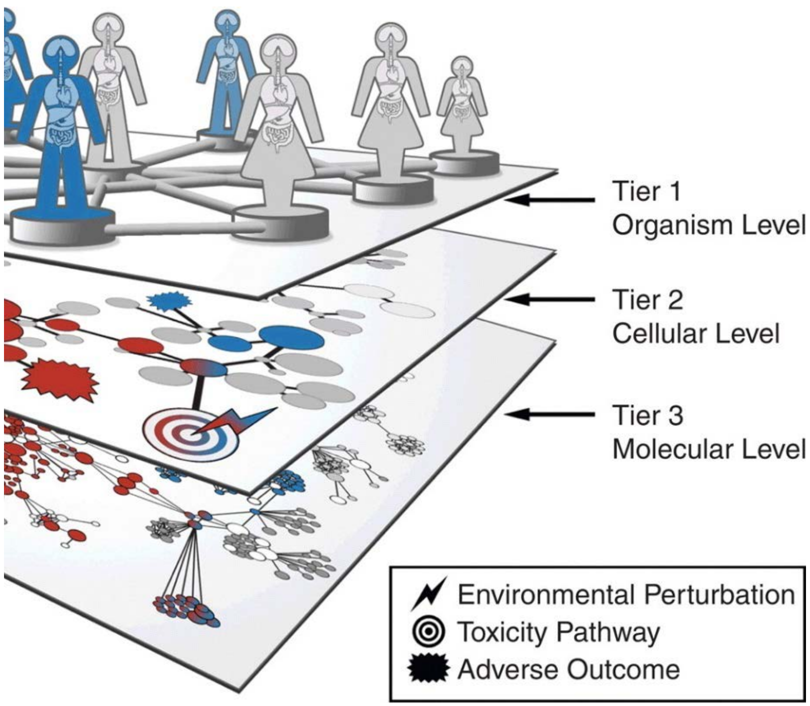


Figure from Kristin Isaacs

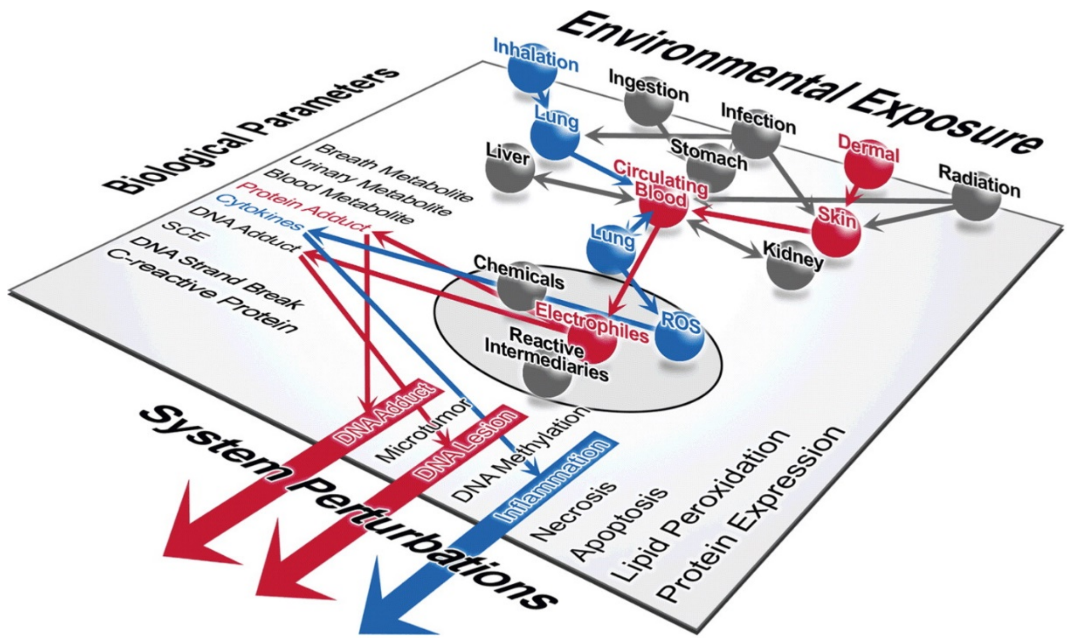
EPA United States Environmental Protection Agency

What Do We Mean By Pathway?



Definition of "pathway" is fuzzy here:

- Not talking about biology
- But human activity and toxicokinetics are both significant factors



Toxicokinetics:

- Inhalation
- Dermal,
- Ingestion

Figures from Pleil and Sheldon (2011)

Knowledge of Exposure Pathways

"In particular, the assumption that 100% of [quantity emitted, applied, or ingested] is being applied to each individual use scenario is a very conservative assumption for many compound / use scenario pairs."

Environmental Science & Technology

Article

pubs.acs.org/est

Risk-Based High-Throughput Chemical Screening and Prioritization using Exposure Models and in Vitro Bioactivity Assays

Hyeong-Moo Shin,[*,†] Alexi Ernstoff,[‡,§] Jon A. Arnot,[‖,⊥,#] Barbara A. Wetmore,[∇] Susan A. Csiszar,[§] Peter Fantke,[‡] Xianming Zhang,[○] Thomas E. McKone,[◆,¶] Olivier Jolliet,[§] and Deborah H. Bennett[†]

[†]Department of Public Health Sciences, University of California, Davis, California 95616, United States

[‡]Quantitative Sustainability Assessment Division, Department of Management Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark

[§]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan 48109, United States

[‖]ARC Arnot Research and Consulting, Toronto, Ontario M4M 1W4 , Canada

[⊥]Department of Physical and Environmental Sciences, University of Toronto, Scarborough, Toronto, Ontario M1C 1A4, Canada

[#]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada

[∇]The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina 27709, United States

[○]Harvard School of Public Health and School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States

[◆]Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720 , United States

[¶]School of Public Health, University of California, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: We present a risk-based high-throughput screening

Potential exposure from exposure

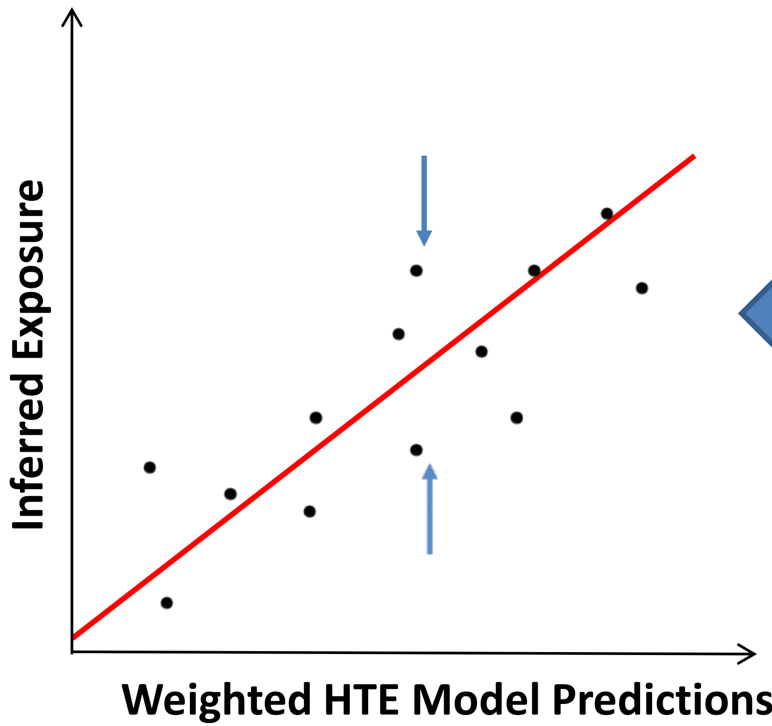
Potential hazard from in vitro

Shin et al., 2015

SEEM is a Linear Regression

__Multiple regression models__:

Log(Parent Exposure) = a + m * log(Model Prediction) + b* Near Field + ε

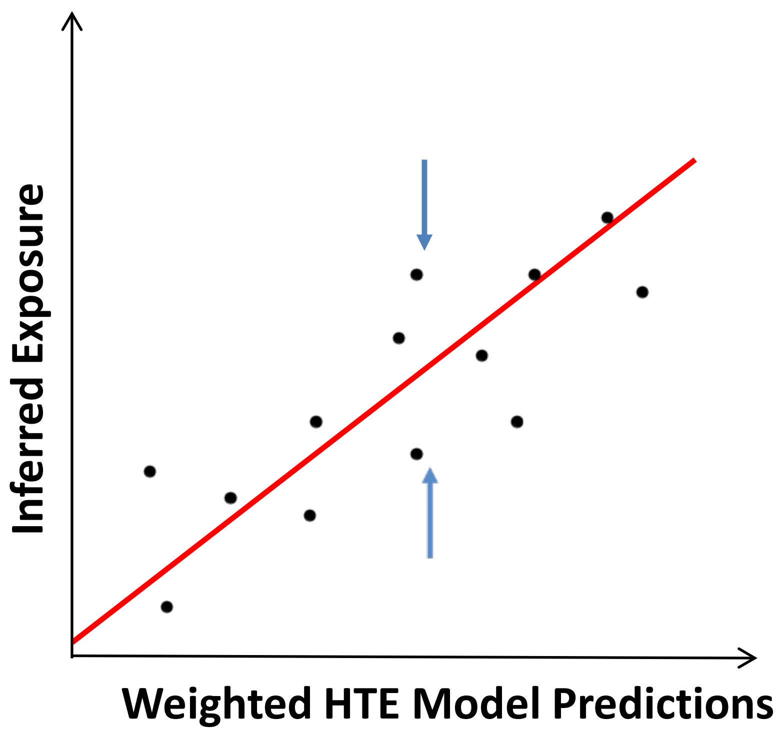


$\varepsilon \sim N(0, \sigma^2)$
Residual error, unexplained by the regression model

SEEM is a Linear Regression

Multiple regression models:

Log(Parent Exposure) = *a* + *m* * log(Model Prediction) + *b** Near Field + *ε*



Not all models have predictions for all chemicals

- We can run SHEDS-HT (Isaacs et al., 2014) for ~2500 chemicals

What do we do for the rest?

- Assign the average value?
- Zero?

Pathway Predictors:
Chemical Use Identifies Relevant Pathways

When averaging over many exposure models, the key is to know which one to use…

Pathway	Positives	Negatives	OOB Error Rate	Positives Error Rate	Balanced Accuracy	Sources of Positives	Sources of Negatives
Dietary	2429	13331	7.8	34	92	FDA CEDI, ACToR USEdb, NHANES Curation	ACToR USEdb, NHANES Curation
Near-Field	1382	3498	20	51	80	CPCPdb, ExpoCast, NHANES Curation	ACToR USEdb, NHANES Curation
Far-Field Pesticide	1726	9204	9.2	48	91	REDs, ACToR USEdb, NHANES Curation	NHANES curation, Diet Positives, ACToR USEdb, NHANES Curation
Far Field Industrial	3183	3792	18	21	82	USGS Water Occurence, ACToR USEdb, NHANES Curation	ACToR USEdb, Dietary and Pesticde Positives

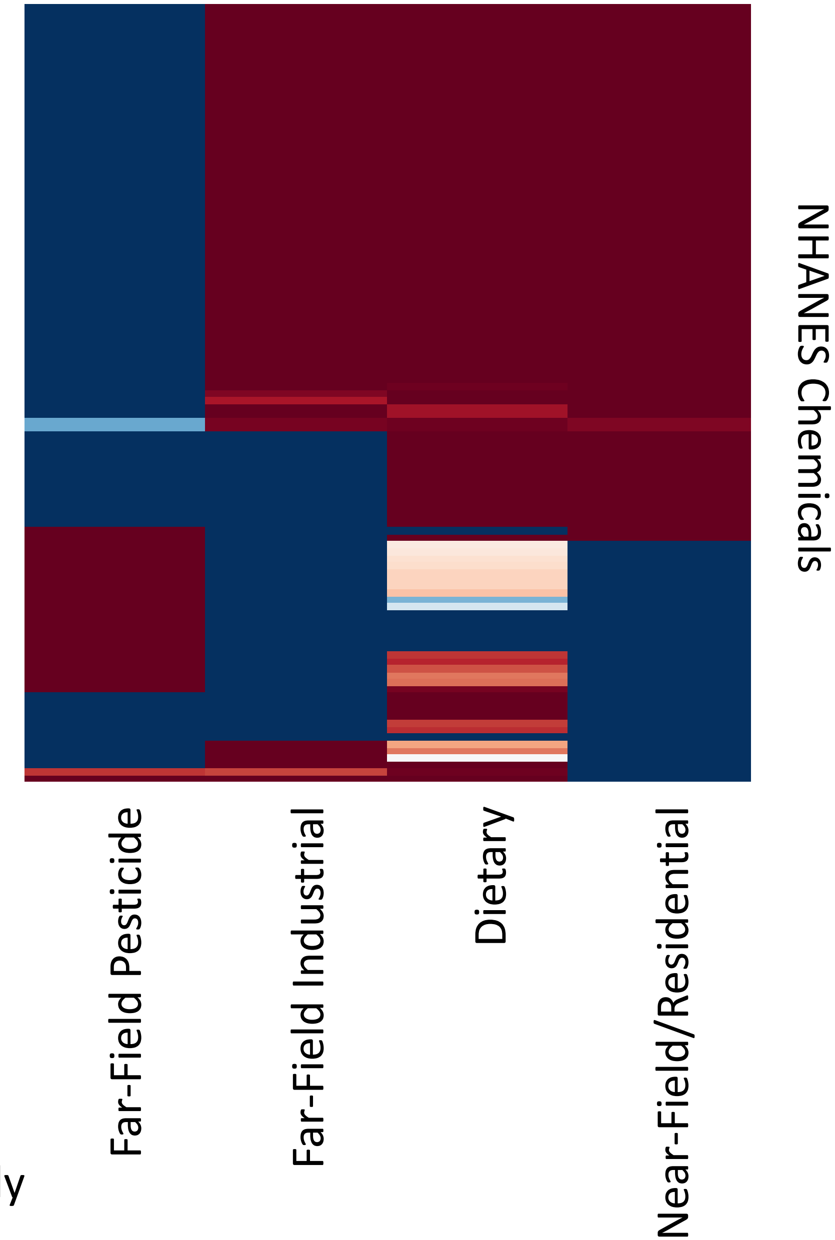
Arbitrary pathway choices
Need better ontology

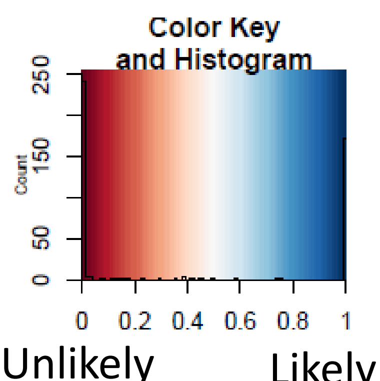
Using Random Forest to predict based upon production volume, OPERA phys-chem, and ToxPrint structure descriptors

EPA United States Environmental Protection Agency

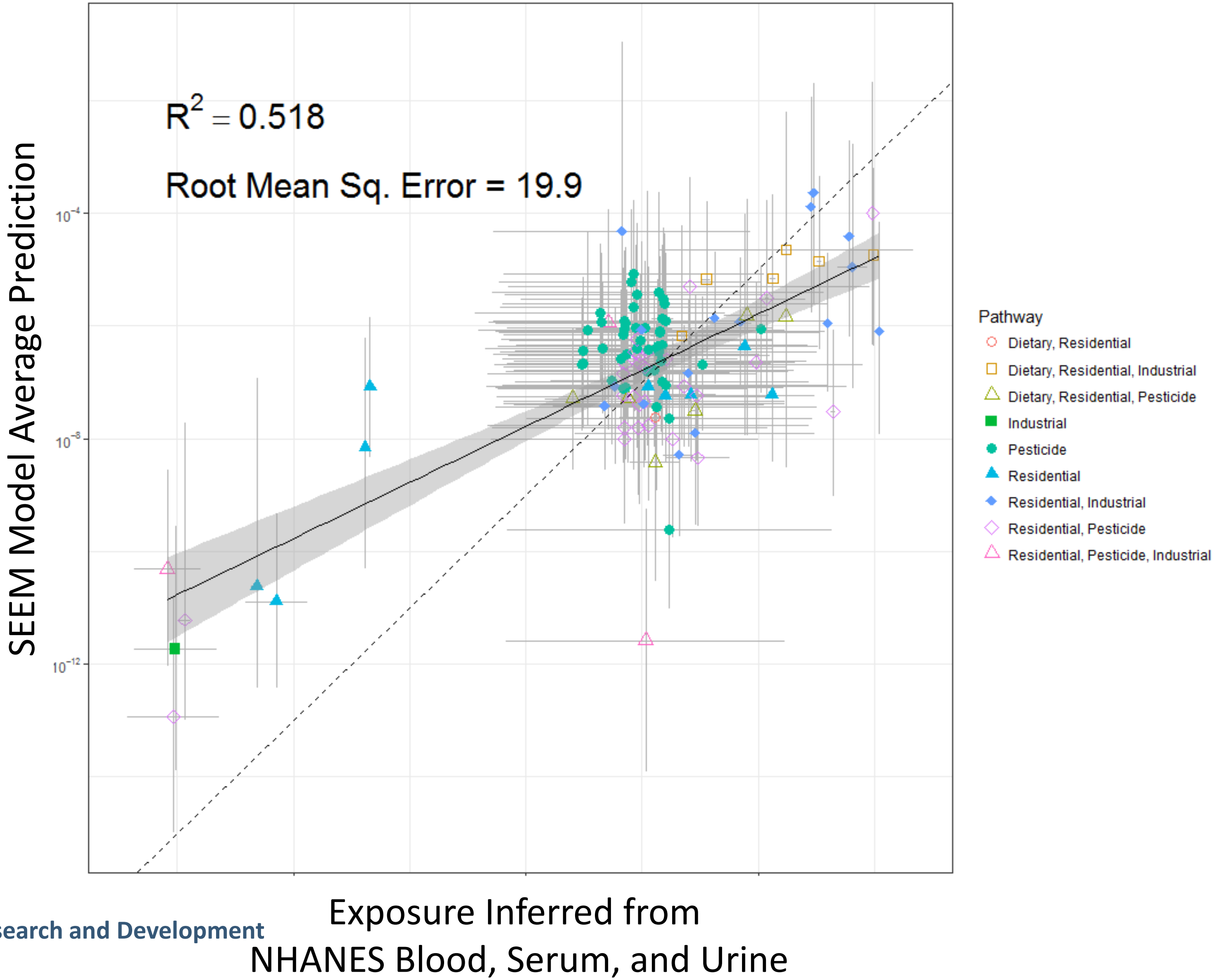
Pathway Probabilities

- Pathways predicted from production volume, OPERA physico-chemical properties and ToxPrint structure descriptors
- Machine learning (Random Forest) – generates a chemical specific probability of exposure by that pathway (used as a Bayesian prior)
- Manual inspection determined that tools we had were pretty lousy for NHANES, so did a manual curation guided by CPcat (Dionisio, 2015)





Third Generation SEEM



United States Environmental Protection Agency

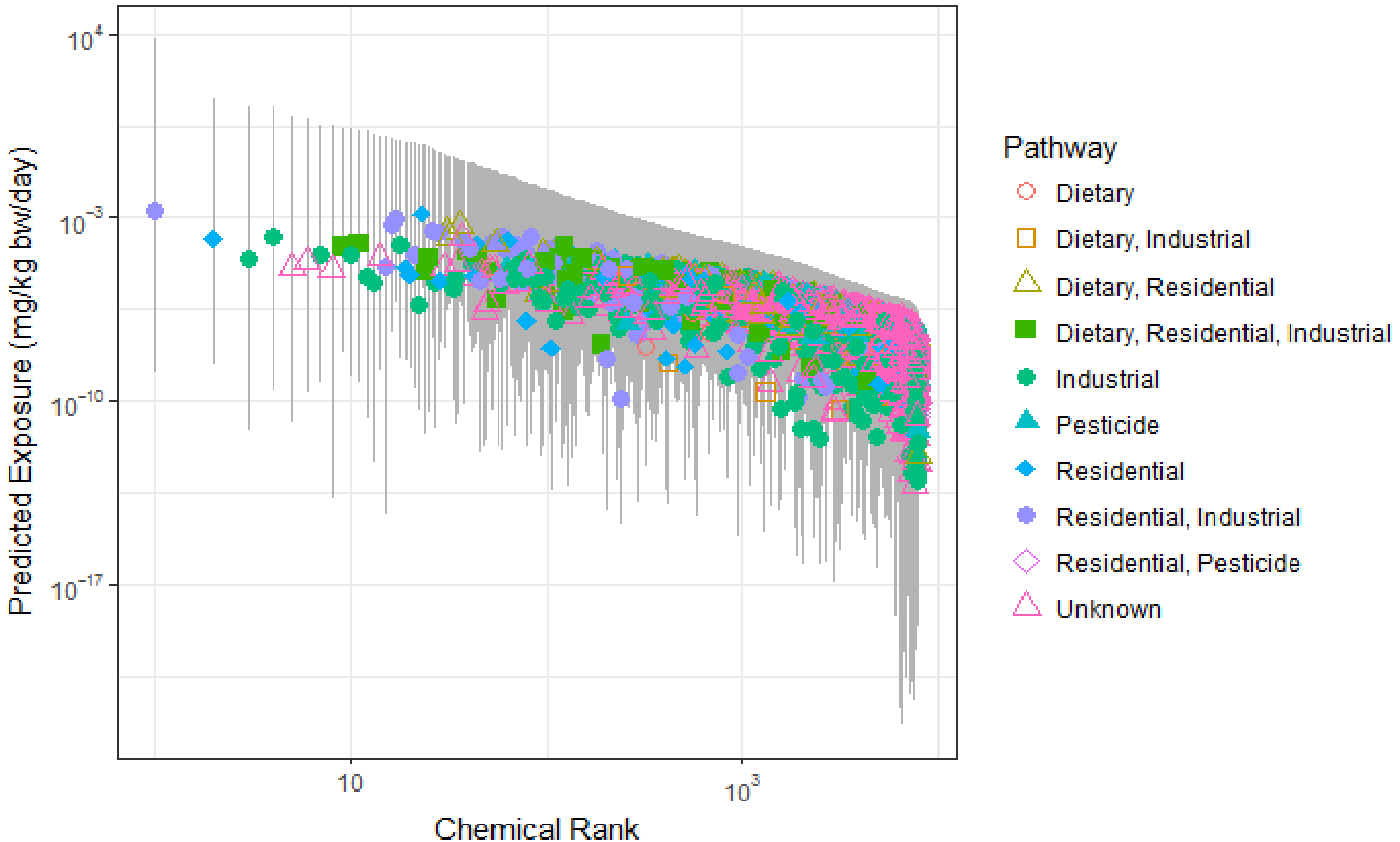
Model Coefficients

	Pathway Mean	NHANES	All Chemicals (Pred.)	SHEDS-HT	Pest Docs	RAIDAR	USETox	Prod. Vol
Grand Mean (Unexplained)	**-15.1 (0.665)**	23	71.50%					
Dietary	-0.0654 (0.213)	6	0.11%	-0.288 (1.13)				1.1 (1.83)
Residential	**0.405 (0.196)**	17	2.03%	**2.15 (0.775)**				**1.36 (0.385)**
Pesticide	**-0.531 (0.113)**	89	12.40%		0.438 (0.671)	0.419 (0.527)	**-4.57 (0.576)**	0.326 (0.846)
Industrial	**-1.77 (1.02)**	2	13.70%			-2.05 (3.13)	-0.808 (1.38)	2.73 (3.01)

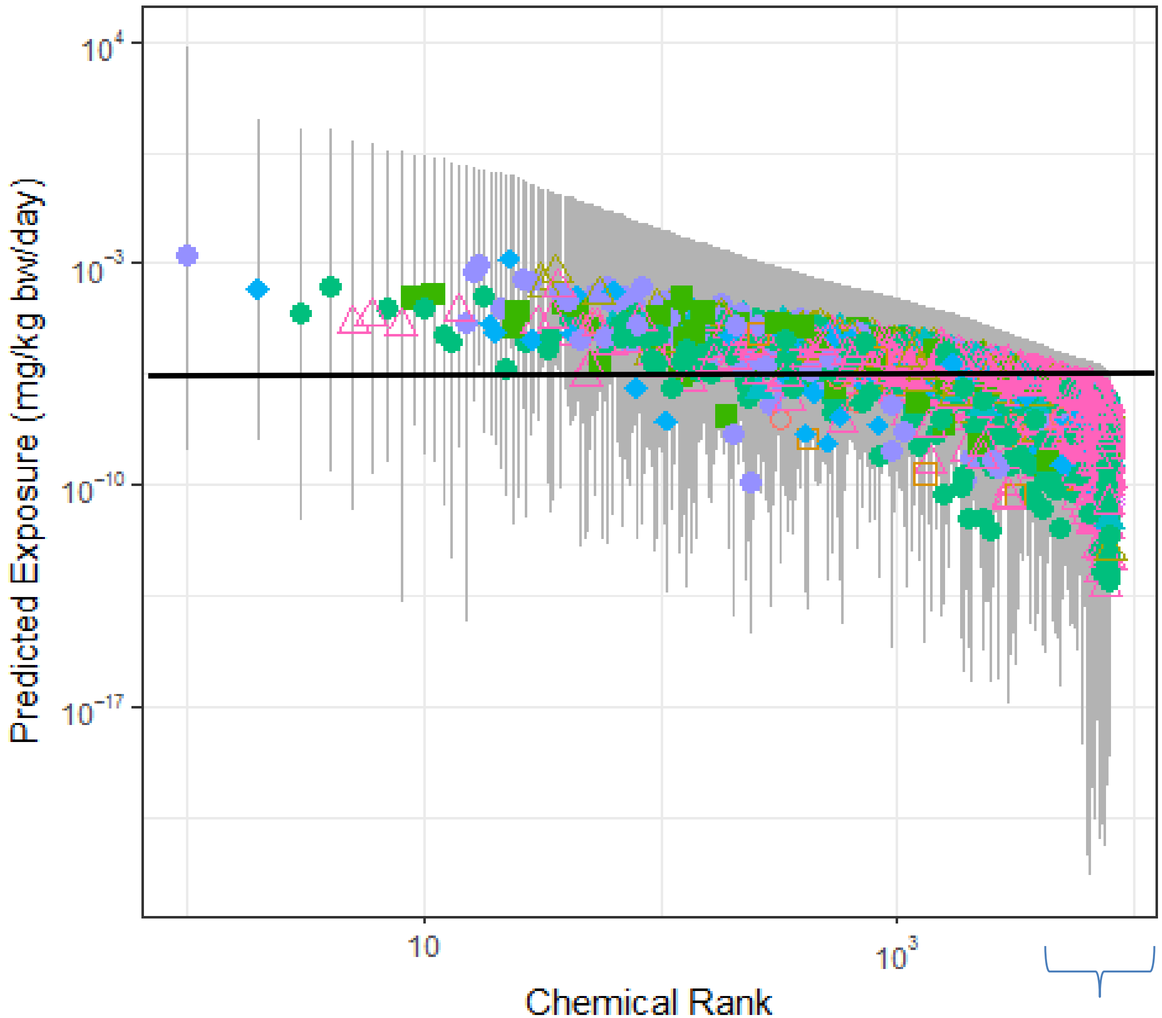
The pathway means recapitulate the Wambaugh et al. (2014) heuristics model (with dietary pathway added)

The significant predictors (mean +- standard deviation beyond zero) are in bold: SHEDS-HT Residential, Production Volume, and USEtox

Human Exposure Predictions for 134,521 Chemicals



Human Exposure Predictions for 134,521 Chemicals

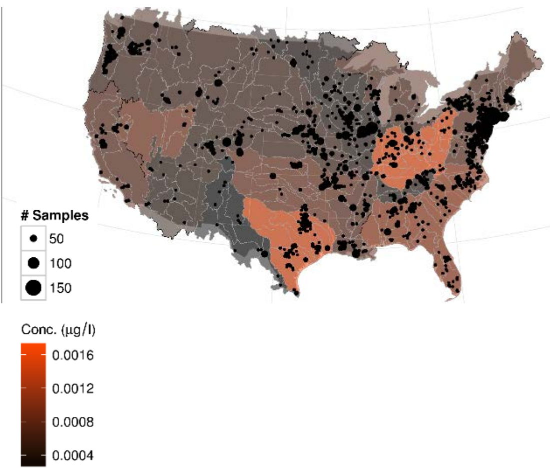


Lowest NHANES limit of detection (LOD) roughly corresponds to ~10^{-6} mg/kg BW/day

95% confident that median population would be <LOD for thousands of chemicals

EPA
United States
Environmental Protection
Agency

Ecological SEEM



Dataset
Setzer et al., (in prep)

Chemicals
n = 91

- **Pesticides**
- **PAHs**
- **Solvents,…**

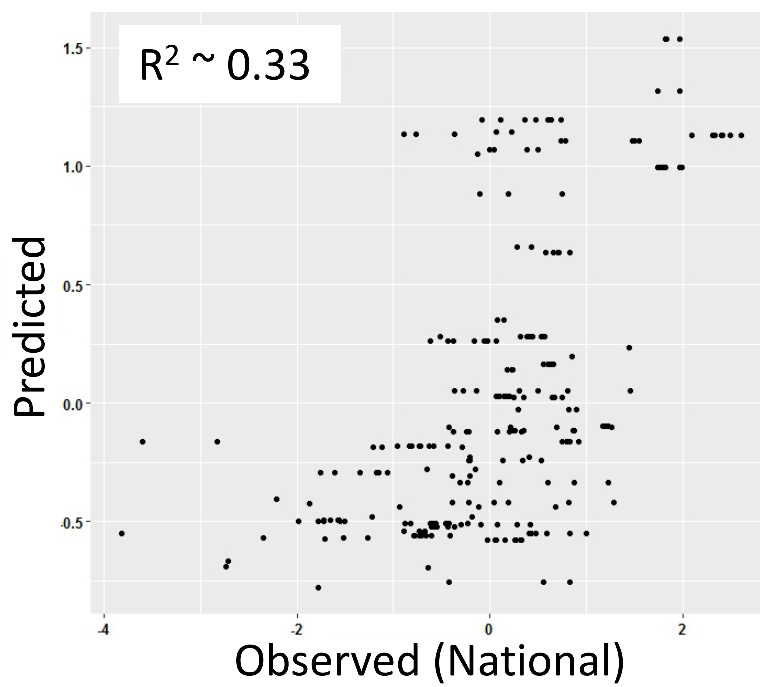
Watersheds

- **1984 – 2014**
  Aggregated by season
- **3 Levels**
  - **National**
  - **Region** (HUC2) n = 18
  - **Sub-region** (HUC4) n = 196

HUC = hydrological unit

Fate and Transport Models

USETox (n = 82)	RAIDAR (n =74)	HT-EXAIR (n = 91)
Rosenbaum et al., 2008	Arnot et al., 2006	Barber et al., 2017



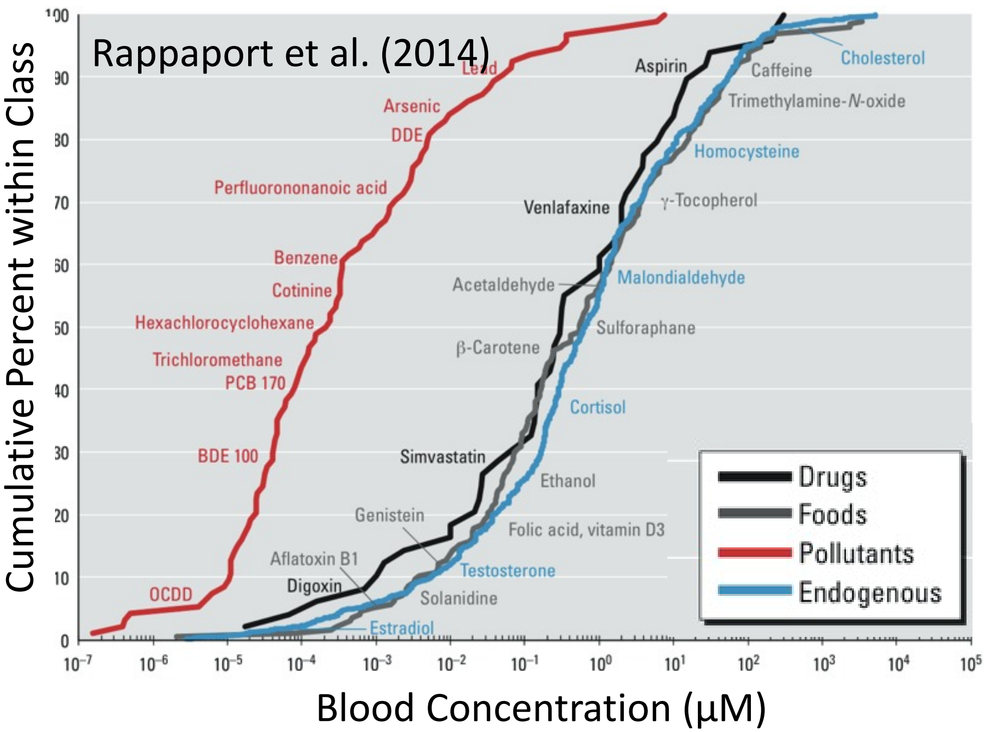
Analysis led by Parichehr Saranjampour

Where Do We Go From Here?

- Models incorporate Knowledge, Assumptions and Data
- The key is to know which model to use and when
- Rough exposure assessments may be potentially useful if the uncertainty can be quantified and is acceptable (i.e., "fit for purpose")

Challenges:

- Using existing chemical data to predict pathways
    - Need better training data for random forest
    - (How do you know something isn't an industrial chemical?)
- Eventually we have got to go beyond NHANES (~100 chemicals)
    - Non-targeted analysis of blood may eventually be possible



United States
Environmental Protection
Agency

Collaborators

Chemical Safety for Sustainability (CSS) Rapid Exposure and Dosimetry (RED) Project

NCCT

Chris Grulke
Greg Honda*
Richard Judson
Andrew McEachran*
Robert Pearce*
Ann Richard
Parichehr Saranjampour*
Risa Sayre*
Woody Setzer
Rusty Thomas
John Wambaugh
Antony Williams

NRMRL

Yirui Liang*
Xiaoyu Liu

NHEERL

Linda Adams
Christopher Ecklund
Marina Evans
Mike Hughes
Jane Ellen Simmons

*Trainees

NERL

Craig Barber
Namdi Brandon*
Peter Egeghy
Jarod Grossman*
Hongtai Huang*
Brandall Ingle*
Kristin Isaacs
Sarah Laughlin-Toth*
Seth Newton
Katherine Phillips
Paul Price
Jeanette Reyes*
Jon Sobus
John Streicher*
Mark Strynar
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
Barbara Wetmore

Lead CSS Matrix Interfaces:

John Kenneke (NERL)
John Cowden (NCCT)

Arnot Research and Consulting
Jon Arnot
Battelle Memorial Institute
Anne Louise Sumner
Anne Gregg
Chemical Computing Group
Rocky Goldsmith
National Institute for Environmental Health Sciences (NIEHS) National Toxicology Program
Mike Devito
Steve Ferguson
Nisha Sipes
Netherlands Organisation for Applied Scientifi Research (TNO)
Sieto Bosgra
Research Triangle Institute
Timothy Fennell
ScitoVation
Harvey Clewell
Chantel Nicolas
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Tox Strategies
Caroline Ring
University of California, Davis
Deborah Bennett
Hyeong-Moo Shin
University of Michigan
Olivier Jolliet
University of North Carolina, Chapel Hill
Alex Tropsha

United States Environmental Protection Agency

References

- Arnot, Jon A., et al. "Screening level risk assessment model for chemical fate and effects in the environment." Environmental science & technology 40.7 (2006): 2316-2323.
- Barber MC, et al. “Developing and applying metamodels of high resolution process-based simulations for high throughput exposure assessment of organic chemicals in riverine ecosystems, “ *submitted*
- Dionisio, Kathie L., et al. "Exploring consumer exposure pathways and patterns of use for chemicals in the environment." Toxicology Reports 2 (2015): 228-237.Rosenbaum, Ralph K., et al. "USEtox—the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment." The International Journal of Life Cycle Assessment 13.7 (2008): 532.
- Isaacs, Kristin K., et al. "SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources." Environmental science & technology 48.21 (2014): 12750-12759.
- Pleil, Joachim D., and Linda S. Sheldon. "Adapting concepts from systems biology to develop systems exposure event networks for exposure science research." Biomarkers 16.2 (2011): 99-105.
- Rappaport, Stephen M., et al. "The blood exposome and its role in discovering causes of disease." Environmental Health Perspectives (Online) 122.8 (2014): 769.,
- Shin, Hyeong-Moo, et al. "Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays." Environmental science & technology 49.11 (2015): 6760-6771.
- Wambaugh, John F., et al. "High-throughput models for exposure-based chemical prioritization in the ExpoCast project." Environmental science & technology 47.15 (2013): 8479-8488.
- Wambaugh, John F., et al. "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals." *Env. science & technology* (2014).
- Wetmore, Barbara A., et al. "Incorporating High-Throughput Exposure Predictions with Dosimetry-Adjusted In Vitro Bioactivity to Inform Chemical Toxicity Testing." *Toxicological Sciences* 148.1 (2015): 121-136.